DOCUMENT RESUME

ED 240 528 CS 007 519

AUTHOR Fry, Edward
TITLE A Theory of Graphs for Reading Comprehension and Writing Communication.
PUB DATE [83]
NOTE 51p.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Viewpoints (120)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Charts; *Classification; Definitions; Elementary Secondary Education; *Graphs; *Nonprint Media; *Reading Comprehension; Reading Instruction; *Theories; Writing (Composition); Writing for Publication; Writing Instruction

ABSTRACT

Graphs are increasingly being used in written communication and writers expect readers to understand them. One definition of graphs--information transmitted by position of point, line or area on a two dimensional surface--excludes displays composed chiefly of numbers or words such as tables or outlines. However, it does include time lines, flow charts, curve (line) graphs, bar graphs, scattergrams, pie graphs, dials, maps, pictures, and abstract drawings because they all transmit information nonsymbolically on a two dimensional surface. Fundamental to comprehending (reading) or writing a graph is knowing the types of underlying coordinates, the scaling, and other factors. The underlying theory is that all graphs are basically one dimensional, two dimensional, or three dimensional. A taxonomy of graphs can help elucidate this theory with numerous categories and examples. Together, the definition, theory, and taxonomy can be useful for teaching and testing graph knowledge in the public schools and in helping writers and readers in many communication media, including journals, textbooks, popular press, computer graphics, and television. (Appended is a taxonomy of graphs.) (Author/HOD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED240528

A THEORY OF GRAPHS FOR READING COMPREHENSION AND WRITING COMMUNICATION

BY

EDWARD FRY
DIRECTOR, READING CENTER
RUTGERS UNIVERSITY
GRADUATE SCHOOL OF EDUCATION
10 SEMINARY PLACE
NEW BRUNSWICK, NEW JERSEY 08903

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Edward Fry

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS007519

ABSTRACT

A THEORY OF GRAPHS FOR READING COMPREHENSION
AND WRITING COMMUNICATION

Graphs are increasingly being used in written communication and writers expect readers to understand them. This article defines graphs in a somewhat unique way as "information transmitted by position of point, line or area on a two dimensional surface." Basically this definition excludes displays composed chiefly of numbers or words such as tables or outlines. However, it does include time lines, flow charts, curve (line) graphs, bar graphs, scattergrams, pie graphs, dials, maps, pictures and abstract drawings because they all transmit information non-symbolically on a 2 dimensional surface. Fundamental to comprehending (reading) or writing a graph is knowing the types of underlying coordinates, the scaling and other factors. The underlying theory is that all graphs are basically one dimensional, two dimensional or three dimensional. A taxonomy further elucidates this theory with numerous categories and examples. The definition, theory and taxonomy are seen as useful for teaching and testing graph knowledge in the public schools and in aiding writers and readers in many communication media including journals, textbooks, popular press, computer graphics, and television.

A THEORY OF GRAPHS FOR READING COMPREHENSION AND WRITING COMMUNICATION

The purpose of this paper is to set forth a theory of graphs[1] for readers and writers. The understanding of graphs is closely related to reading comprehension of prose. Graphs in many forms such as bar graphs, time lines, maps, and pictorial illustrations are an increasingly used form of communication in books, magazines, and newspapers. Graphs are also an important form of visual communication in the electronic medium such as the cathode ray tube in television sets or computer monitors. Tufte (1983) estimates that there are between 900 billion and 2 trillion statistical graphs printed in the world each year.

Nearly anyone's definition of a good reader would include the ability to comprehend a written paragraph. I would like to add that it is increasingly important that the good reader be able to comprehend all forms of graphs.

The obverse side of reading is writing, and I would further like to suggest that writers of all types should be better able to understand and use a wide variety of graphs in communicating with the reader.

Writing and to some extent reading, are arts which allow for individual differences. There are wide varieties

---

[1]This term is not synonymous with "Graph Theory" as used by mathematicians. This theory of graphs is more related to schema theory, language comprehension, artificial intelligence, statistical data displays, cartography, and art.

of vocabularies and styles which can be used in writing with words and few would argue against the idea that the writer and the readers should have a common verbal vocabulary. But the "vocabulary" of graphs is seldom taught, and the underlying structure of a graph is seldom explained in the curriculum of today's schools.

This article then, is aimed at those who teach reading comprehension, at all levels and in all subjects, in the educational system. Secondly, it is aimed at writers, of all types, including writers of textbooks, newspapers, magazines, company reports, and technical articles. Also, it is aimed at writers of the important newer visual media of television and computer programs. Finally, this article might make a modest contribution to the small but important literature on graphs which spans such diverse fields as statistics, social sciences, engineering, cartography, and art.

Defining and Understanding a Graph

It is both simple and complex to define a graph. It is simple to give a one sentence definition and it is complex to see it explained in a taxonomy which covers everything from a time line, to a bar graph, to a political map, to a photograph. But let us begin: A graph is information transmitted by position of point, line or area on a two dimensional surface.

Implicit in this definition is that it is the location in area (2 dimensions) that is the important factor. Excluded are symbols or arrays that are basically symbols. Hence a paragraph of prose or a table of numbers are not graphs because they are primarily composed of symbols. It is, of course, permissible to have some symbols on a graph; a map can have names of countries, for example, or a frequency polygon can have some numbers along the base line, but the primary transmission of information is by position, not symbolic, means.

You might note that the definition states that a graph must use a two dimensional surface such as a piece of paper or a cathode ray tube. Three dimensional models such as statues are not graphs.

The term "chart" and "graph" are roughly synonymous except that "chart" is used in some settings like chart of chemical elements or marine chart (map).

Fundamental to understanding graphs is to realize that they all have an underlying system of coordinates. For example, the common bar graph is really part of a grid (lines at right angles) and that grid is part of the Cartesian coordinate system which gives meaning to each location. The common bar graph uses the upper right hand quadrant of the two dimensional Cartesian coordinate system

but it is quite possible to use other quadrants which use negative numbers. See Figure 1.

[Insert Figure 1 About Here]

This illustration shows the grid and coordinates that underlie most common bar and line graphs. Like most typical graphs, time increases along the +X axis, and amount increases along the +Y axis. Most graphs use only the +Y and +X or the upper right quadrant of the Cartesian coordinate system. Slightly atypical is the use of the -Y axis which shows that on Day 3 Maria lost $5 in operating her lemonade stand and a further atypical use of the -X axis which shows that on the day before opening she lost (or spent) $10. The dotted line shows that the same information can be transmitted by a line or curve (a frequency polygon). The dots on each bar show that the same information could be transmitted by simply placing dots (a scattergram) on the Cartesian coordinate plane. It is the writer's option to use bar, curve or dot, and the readers should be able to read all three with equal comprehension.

A second fundamental in understanding a graph is to know what scaling is used for the coordinate lines. In other words, "What do the grid lines represent?" They can represent an astonishing number of things. For example, a Y axis showing amount can be amount of dollars, tons of wheat, man-hours, or speed in light years. Scaling can be

nominal (non-ranked) like England, France and Germany, or ordinal (ranked) like January, February and March, or interval like 1925, 1930 and 1935 where there is some system of scaling units. Scaling units can have much variety, a scale of numbers can be arithmetic or logarithmetic. The numbers can be conventional (base 10) counting numbers or can be minutes and degrees in which minutes can only go up to 60. In fact, scaling is so important that we will have a more thorough discussion of it later.

In summary then, a graph transmits information by position or area on a two dimension surface, and fundamental to understanding any graph is to know what coordinate system is being used and what scaling is used on each axis. However, it should also be mentioned that graphs transmit some information by use of some symbols such as words for titles, numerals for axes, color or shading, and use of some special symbols such as arrows to indicate motion.

Coordinates

Rene Descartes is reputed to have been a late riser and to have gotten the idea of using coordinates to describe points and lines in space by watching a fly walk around on the ceiling. Coordinates have been used earlier for maps but this brilliant French mathematician began using algebraic notations like 2x-3y to describe a point, two points to describe a line, and formulas to predict or

locate any point on a complex curve. To this day our rectangular system of coordinates is named in his honor. A brief illustration of his two dimensional coordinate system is given in Figure 1. A third dimension, of Z axis, can be used to describe any point in space (technically area is 2-D and space is 3-D). For example, the third dimension is used regularly to show the heights of mountains on topolographical maps. The view shown on common topolographical maps is as though looking down from straight overhead, but some maps have a perspective as though looking at the land from a plan at a 45 degree angle. This view is similar to a photograph taken from the side window of an airplane. Some maps, many schematic drawings, and most photographs represent points (information) in three dimensional space, and any point in three dimensional space can be described using Cartesian coordinates with X, Y, and Z axes.

There are times, however, when it might not be useful to use the rectangular Cartesian coordinates. The next most common system of coordinates is the polar coordinate system. In this system there is a dot in the center of concentric circles and radiating lines. See Figure 2. Polar coordinates were first widely used by cartographers to get rid of some of the distortion in polar regions of the earth which resulted when the typical Mercator projections

were used. However, most readers will be most familiar with it as the underlying coordinate system in pie charts or the traditional round clock face. Polar coordinates have many other uses in engineering and scientific graphs.

[Insert Figure 2 About Here]

Some coordinate systems have special purpose uses like the Latitude and Longitude coordinate system used to describe locations on the earth's spherical surface. Others describe curved space that some mathematicians theorize about.

Scaling

Scaling is an extremely important factor in understanding graphs. First of all you cannot begin to read or write a graph without knowing, or at least making assumptions about, the kind of scaling used. We might begin by reviewing a description of the basic kinds of scaling such as might be found in an opening chapter of most statistics books.

Nominal scaling is simply the use of relatively unrelated names. For example, if we wish to make a bar graph of wheat production in Europe we might label the X axis France, England and Germany. In nominal scaling there is no implied ranking or inherent order. This is perhaps the least "mathematical" of the scales but it allows us great flexibility in the kind of information we can place on a graph.

Ordinal scaling is the use of ranked information or data.

For example, the months are definitely ranked: January, February, March, etc. March never comes directly after January or before February. This particular set of ranked data, the 12 months, is also cyclical; right after the last month, December, starts the beginning one, January. Hence this scale might sometimes be most appropriate on a circular base line of a polar coordinate graph. You might also note that ordinal scaling is not concerned with interval or unit size or regularity - one month has 28 days and some have 31 days. A common example of ordinal scaling would be Step 1, Step 2, Step 3, etc. that we see in the flow chart where direction and order (ranking) are of great importance but size, kind, or distance of step are relatively unimportant for this scaling.

Interval scaling has data which is not only ranked as in ordinal scaling but each step is in some kind of regular unit. There is a definite and usually numerical or mathematical relationship between each unit on the scale. Some examples of scaling units for interval scaling are arithmetic numbers (counting), logarithmetic numbers (exponents), percentages, standard scores, or the degrees and minutes of latitude and longitude. A more thorough discussion of scaling units will occur in the next section.

Ratio scaling is not a really different form of scaling but rather it is interval scaling in which the axes start

at zero. An interval scale on the X axis of a bar graph could start at 1925 rather than at zero because it would make the chart rather unwieldy and besides that the data only covers years beginning with 1925. Hence, starting a scale where the variations occur tends to emphasize the variations. For example, if a Dow Jones Stock Market average used a Ratio Scale one would quickly see that the daily fluctuations were only a small percent of the total price of stock. However, you would have a difficult time telling if the stock went up or down between most days or weeks. The good side of ratio scaling is that it emphasizes differences so we can see them; the bad side of ratio scaling is that it tends to hide or down play the 0 or total amount being shown. See Figure 3.

[Insert Figure 3 About Here]

Scaling Units

Both the interval scales and the ratio scales which we have been discussing have some regular or mathematical relationship between the units or lines on the coordinate grid. Any coordinate system can have any of the following kinds of units:

Arithmetical Units are by far the most common type of units; these are part of the number system we regularly use to show amount. It becomes important to understand this particularly when we discuss logarithmic units later or

some other unit systems, such as time or spherical degrees. Arithmetical numbers like any number can stand for many things; 250 can be dollars, tons of wheat, test score points, etc. It is, of course, absolutely essential to know what the numbers along the axes mean before reading the graph.

Percentage Units are useful in simplifying and explaining some phenomena. They are particularly useful if you are comparing two kinds of units. For example, if you wish to compare car prices with the average wage paid an auto worker over a 10-year period you might have two curves on the chart with the Y axis in percent.

Standard Score Units are based on the standard deviation of a normal distribution curve. A good explanation of them can be found in any elementary statistics book, but briefly standard scores have a mid-point with the standard score moving out from the mid-point in regular standard deviation units. For example, the historic mid-point of the kind of standard scores used for Scholastic Aptitude Tests is 500 with one standard deviation being worth 100 standard score points so 600 would be one standard deviation above the mean. Other tests use other mid-points and other sized standard score units. Any standard score can be translated into a percentile, which tells you where you would rank out of 100 randomly selected students. Note: percentile is based on a population and it is not the same

as a percentage mentioned earlier. Some graphs use a probability scale (log square), which is really a normal distribution curve flattened out so that by plotting a few points and drawing a straight line through them and extending the line, you can estimate the standard score, percentile, or probability (probable frequency in the population) given one input variable, such as a test raw score.

[Insert Figure 4 About Here]

Time and Degree Units are interesting in that they do not follow the regular decimal system. For example, minutes only go to 60 so 75 minutes is one hour and 15 minutes. Days count up to seven before becoming a week, etc. Angular distance, such as latitude, have minutes and seconds, which, like time, minutes and seconds only count up to 60, and degrees only count up to 360.

Logarithmic Units are a somewhat sophisticated type of unit which are very valuable in certain types of situations, but they require some knowledge on the part of the reader and writer. A logarithm or log is an exponent usually to the base 10 so that 100 can be expressed as 10^2. In this case, two is the exponent or in other words, the log of 100 is two. Any number can be expressed in logs, for example, five is .699 and three is .477. Many mathematics texts can give you a full explanation of

the properties of logarithms, but for graph purposes, using a log scale does some interesting things. A log scale shows percentage change not arithmetic change. If one inch represents the distance between two and four on an arithmetic scale, then the distance between four and eight will be four inches, but using a log scale, if one inch represents the distance between two and four, then the distance between four and eight will also be one inch because the change between two and four and the change between four and eight are both 100%. This can cause lines which appear to be straight on an arithmetic scale to curve when a logarithmic scale is used. See Figure 5. Many graphs which use a logarithmic scale on one axis use an arithmetic scale on the other axis as we did in Figure 5. This type of graph is called semilogarithmic. Log scales can be used when two different values are used like the percentage scales mentioned earlier. They also can be used when very large number differences occur. For example, a change in scale value from 10 to 100,000 would set up a lot of space on an arithmetic scale, but could be compression on a logarithmic scale with several cycles. Log scales can also tend to distort curves at the upper or lower end as compared to arithmetic scales. An important part of log scaling is that it can be used to analyze data

by casting it in a different visual image.

[Insert Figure 5 About Here]

Averaging Units tend to smooth out small fluctuations. If you are having trouble seeing if the school district's seventh grade classes test scores follow a normal distribution curve plot only deciles, or group the students in groups of 10 or 6 and plot the average of the group. Some scientific and financial graphs use a moving average which means they plot an average over some time period. For example, daily fluctuations in a stock might obscure a downward trend that would be visible if a moving average of five days was used (each day you plot the average of the previous five days). For reading a graph it is important to know if raw data or averages are used.

Perspective Scaling Units are important in understanding pictures and many 3-D graphs. A perspective scale can easily be visualized by a photograph of railroad tracks in which the ties to get closer and closer together as they recede from the viewer towards a vanishing point. In a photograph, a man 100 feet away is half the size on the photograph of a man at 50 feet. Mathematically, a perspective scale is like a tangent (side opposite/side adjacent). If you double the distance (side adjacent) you must double the size (diameter or side opposite) of the object for it to appear the same size on the graph

(picture). Another way of looking at perspective scaling units is that they are reciprocal numbers; one divided by one equals one, one divided by two equals.5, one divided by three equals .33, one divided by four equals .25, etc. Thus, a man who is photographed one unit away, say ten feet, will be only one third as large on the photograph if he is three units or 30 feet away. We will discuss pictures and 3-D graphs later, at this point it is only important to realize that one type of regular scaling is perspective scaling. A few bar charts would use perspective scaling to glamorize them or emphasize certain aspects, but this is often frowned upon because perspective scaling tends to distort statistical data.

Variable Scaling Units are not often used. This means that there is an irregular relationship between different parts of the scale. It is useful from a theoretical standpoint of explaining cartoons or abstract art where there is intentional distortion. It is very unusual or perhaps cheating to use variable scaling in most interval scale graphs.

Other Scaling Units are possible, such as square roots, radii, and other specialized units. These tend to occur in engineering and scientific fields and are not frequently seen by the non-scientific reader. However, they tend to point out the importance of knowing the kind of units

used on any graph's scaling.

A Theory of Graphs

Now, having laid the groundwork with a definition of graphs, plus a discussion of coordinates and scaling, I would now like to propose a basic theory of graphs and later follow it up with a taxonomy and discussion. My theory of graphs is that all graphs are basically one dimensional, two dimensional, or three dimensional information presented on a two dimensional surface.

A one dimensional graph might be visualized as a thermometer or a time line. In a thermometer we have a numbered scale in degrees and the information is simply the point at which the mercury is resting. In a simple time line, we have a double scale; we might have years on one side of the scale (interval scale) and events along the other side (nominal scaling). It is the position (location) of the event along the years scale which conveys the information and clearly shows by space its relationship in time to other events. For illustration of these and other graphs see the Taxonomy section.

More complex one dimensional graphs would include branching, such as a flow chart. Flow charts are one dimensional because the basic information is lineal. It is most important that Step one follow Step two, but it is of little importance (no information) whether it is put

towards the right or towards the left on a two dimensional piece of paper.

From a human visual point of view it is also readily apparent that a time line or a flow chart looks quite different from other kinds of graphs, such as a bar graph or a pie graph.

Two dimensional graphs are the most common kind. Nearly everyone would agree that a bar graph is a graph. The most common coordinates are the Cartesian rectangular coordinates and as we have discussed, there is a wide variety of scaling. Persons who like a narrow definition of graphs would also restrict their use to some kind of numerical data though obviously a nominal versus a nominal scale could easily produce a graph with no numbers. For an example, see the status graph in the Taxonomy.

The familiar music notation system which had its origins in the 12th century is a two dimensional graph with time progressing along the X axis and sound frequency increasing up the Y axis.

There is little less agreement on whether a map is a graph. I would definitely say that maps are graphs and the following taxonomy should help to make this point. Basically, maps transmit information by space. They have dots, lines and areas superimposed on a coordinate system. The fundamental information a map transmits is the

location of that dot, line, or area and its relationship again in space to another dot, line or area.

Three dimensicnal graphs would typically be pictures or maps which show a third dimension. It is a difficult and unusual use of the word "graph" to state that a picture is a graph, but bear with me for a moment. If you take an aerial photograph of your city from straight overhead, would it not look quite similar to a traditional map of your city? Both would have the same theoretical underlying coordinate system and both would have similar information such as street intersection location. The big difference would be that the photograph would have more information or "pixels" as those in computer science would say.

The concept of pixel is quite fundamental to understanding the relationship between a picture and a traditional map or in understanding why a picture is a graph. A pixel can be viewed as one bit of information; it is either off or on. Your T.V. screen or any photograph is actually composed of many tiny dots which are either on and lighted or dark (photographers call this "grain"). The greater the number of pixels, the better the resolution or sharpness of the picture.

Now, if instead of photographing the town through the floor of the airplane, we photograph it out the side

window at a 45 degree angle to the earth, do we suddenly have no spatial information? Can we no longer see the relationship between Street A and Street B? The answer is that what we have done is to introduce a third dimension (we can now see how tall the buildings are) and we have altered the scaling to a type of perspective scaling (the distant parallel streets seem to converge). I could now take this photograph to an artist and ask him to draw a picture of the town and he might choose to leave out some of the details (pixels) but he could make it look either 2-D or 3-D. In any event, the artist's picture, the photograph and the town map would all have a theoretical system of coordinates underlying them and like all coordinates, they would have scaling. Furthermore, like all graphs, they would transmit information by location on a two dimensional surface.

Now it is but a small step further to see that the photograph of the town taken from the side of the plane so that we can see the height of the building is not fundamentally unlike a photograph of a person or a flower. A three dimensional system of coordinates with appropriate scaling can correctly identify where the pixels should fall on the two dimensional photographer's negative or artist's canvas. Hence, all pictures are graphs.

A Taxonomy of Graphs

A taxonomy is a scientific classification system.

The best known one is certainly the classification of plants and animals used in the biological sciences. Taxonomies in education and psychology are considerably less exact or verifiable, nevertheless, they are useful in certain understandings and applications. Bloom's Taxonomy has been useful in stimulating the development of tests and curriculum materials. It is hoped that this Taxonomy of Graphs might serve a similar useful purpose.

The Taxonomy of Graphs is based on the foregoing definition and theory that all graphs transmit spatial information and have an underlying structure of dimensionality, coordinates and scaling. However, the taxonomy must deal with the real world of graph makers. It should be able to accurately or at least approximately classify all graphs which appear in most textbook, magazines, or other visual media.

One definite limitation of the taxonomy is that it does not attempt to be highly technical in the fields of science, mathematics or engineering. For example, radio engineers use circular graphs which have impedance coordinates. However, many of the graphs in this taxonomy have wide use in many scientific fields and the classification and scalings could be extended into more technical areas at a later time.

In dealing with the real world of graph use, the

taxonomy must at times make some awkward compromises. But, despite this and other limitations, the taxonomy can still classify most graphs in wide use and many that have rather restricted use. Certainly it is an expansion of the definition of scope of "graph" that is found in most textbooks and dictionaries.

An outline of the Taxonomy appears in Figure 6 and the full Taxonomy appears in the Appendix.

[Insert Figure 6 About Here]

Pictures

Of all the sections of this taxonomy and theory that are apt to cause the most consternation and be least accepted at first glance, it is the picture section. Many people will not readily concede that a picture is a graph. To them a bar graph is a typical graph and it obviously uses numbers. However, the new generation is growing up with computers and will have little difficulty in accepting the pixel location concept. In some computer programming courses, students are taught formulas for perspective scaling so that any shape, such as an octagon, can be made to rotate on any axis. The octagon then appears on the screen to have the nearer side longer and the farthest side shorter. More complex programs can rotate whole space ships or automobiles. This clearly demonstrates the regularity of perspective scaling.

Cartoons and schematic drawings are simply pictures with part of the information missing. The many pixels of realism are modified or eliminated so that just bold lines show. The change in form and line in some cartoons is akin to abstract art.

Abstract art introduces yet another variable. In addition to omitting much realistic pixel information the artist has also warped the coordinate grid and varied the scale. In a drawing of a person, the head might appear very large, the body tiny, and the feet large. The artist has simply altered reality by choosing his own grid and scaling units. The viewer now has the problem of trying to see or understand what the artist has done and perhaps why. The further away from visual reality (a regular scale and many pixels), the more difficult it is for the average viewer to see or sometimes to accept the image.

Part of the problem of abstract art lies in what mathematicians call "mapping." Mapping is the correspondence between two sets of points. In a realistic picture, such as a photograph, there is a regular and systematic correspondence between the visual reality and the picture. In abstract art, there is incomplete and distorted mapping or coordination between points. The artist is filtering reality through his mind to communicate perhaps a different message.

Series drawings, such as the evolution of a tadpole, or a series of frames in a typical comic strip, are a method to get onto two dimensional surfaces, a concept of a fourth dimension - time.

Discussion

The popularly accepted father of our most popular statistical graphs was an Englishman named William Playfair who used the bar graph in 1786 and the pie chart in 1801. He also used the curve chart and the surface chart. Prior to that, a Frenchman, Phillippe Bauche used the contour map in 1752 and the historical time line was used by an Englishman Priestly in 1765. However each of these forms has earlier antecedents (Costigan, 1983).

However, as Funkhouser (1937) the historian of this chronology has pointed out, there is little agreement on graph standards. It is not for lack of trying, however, as such respected scientists as Edward L. Thorndike wrestled with the problem in 1915 as a member of the American Statistical Association Graphical Standards Committee.

The situation has not improved lately either. Schmid (1978) states that The American Statistical Association disbanded the committee in 1954 though some work on graph standards did continue by the American National

Standards Institute. Schmid feels that it will be several decades before appreciable progress can be made. One of the few agreements seems to be that the most common type of graph, the time series, should have time increase along the +X axis.

Since there does not seem to be any general agreement on graphing standards, it makes it difficult for the reader to achieve what Samuels and LeBerge (1974) would call "automaticity" or more or less instantaneous reading or comprehension of a graph.

There is something less than agreement on just what constitutes a graph and this confusion is further compounded when the term chart and graph are used either interchangeably by some authors or specifically by other authors. The variety of what can be called a graph also has much variability. For example:

> Graphic presentation represents a highly developed body of techniques for elucidating, interpreting, and analyzing numerical facts by means of points, lines, areas, and other geometric forms and symbols.
>
> \- - -
>
> The term chart, as used in this book, includes a wide variety of techniques for presenting information, such as maps, blueprints, schematics, diagrams, tables, illustrations, graphs and other outline representation. . . (Schmid, 1979)

> Graphs usually take the shape of rectangle with quantative scales shown on the horizontal and vertical axes. (Meyers, 1970)

Twyman (1966) has a "schema (matrix) that attempts to embrace all graphic language." This has a heavy emphasis on symbols, such as words and numbers in lineal and list forms, but his schema would also include maps and pictures.

Jacques Bertin (1980) analyzes graphical methodology with a strong cartographic emphasis. He states that "Every graph and every map, no matter how large is the transcription of a two dimension table." He does not include verbal material as does Twyman, nor does he include lineal or pictorial material as graphs.

MacDonald-Ross (1977) suggests a classification based on purpose; he has sections on algorithms, cartography, comic strips and typography.

This taxonomy, while broader than some, has an underlying definition that graphs are based on presenting basically non-symbolic information and that meaning is derived from spatial location. To understand this, it is necessary to know the coordinate system used and the scaling units.

The taxonomy in this paper is in harmony with Doblin's (1980) Matrix of Information Messages. He has three broad

categories of Verbal, Numerical and Visual messages. This taxonomy is in the Visual category. But his Visual category is further subdivided into three categories of Ideogrammatic, Diagramatic, and Isogrammatic (pictures and maps). My taxonomy is an expansion of Doblin's Diagramatic and Isogrammatic categories. My taxonomy intentionally omits Doblin's category of Ideogrammatic which is largely visual symbols such as Chinese characters or the red octagon stop sign used in the United States. This also explains why my taxonomy leaves out data display systems such as Bachi's Graphical Rational Patterns or Chernoff's Faces. Doblin is also quite concerned with the psychological or communicative impact of messages which is beyond the scope of this taxonomy.

All taxonomies have an emphasis or area of specialization. This taxonomy has relative emphasis and finer classifications in the traditional graph areas of line and bar graphs. Conversely it has a lighter emphasis or differentiation in the areas of maps (cartography) and pictures. These areas could be filled out at a later time without disturbing the overall outline or theory.

It should be pointed out that this taxonomy has relatively pure types of graphs in contrast with graphs found in some publications which use two or more types of graphs mixed together. For example, regional data might

be displayed as a map of the United States which has pie charts superimposed on various regions. There are many types of combinations but the overwhelming majority of combination graphs use graph elements incorporated in this taxonomy.

Part of any good definition of a concept is to describe what is included and what is excluded. This definition of Graphs includes the graphs in the taxonomy which is in the appendix. Excluded are arrays of symbols like letters and numbers used in prose passages, words used in special arrangements like advertisements or numbers in tables. Tables are included as graphs by some graph scholars and I admit that numbers are given meaning by position in the table but tables are excluded from this theory and taxonomy because tables are primarily composed of symbols.

The last section of the taxonomy includes "factors" of graphs such as color and scaling. These factors contribute to the reading and writing of graphs but they are not themselves graphs. This section was included in the taxonomy to remind teachers and writers of these factors. The last section is also for use by persons who may wish to use the factors in classifying graphs.

Conclusion

This theory of graphs and taxonomy have as their main goal the aiding of readers and writers to deal with this

basically non-symbolic medium of communication. It is also hoped that they might have implications for teaching and testing in the public schools. Another important goal in graphing data is to help with the analysis of data, or as it might be put, "seeing what the data mean." Finally, as Wainer and Thissen (1981) so aptly put it after their excellent literature review, this graph theory and taxonomy might "help to support the notion that research on graphics is a proper pursuit for a scholar, and is not merely the drawing of pictures."

REFERENCES

Bertin, J. The basic test of a graph: A matrix theory of graph construction and cartography. In Kolers, Worldstad, & Bouma (Eds.), Processing visible language, 2. New York & London: Plenum Press, 1980.

Biderman, A. Discussion: Innovation, standardization, and testing. In Statistical graphics. Graphic Presentation of Statistical Information, U. S. Census Bureau Technical Paper No. 43. Washington, D.C.: U.S. Dept. of Commerce, 1978.

Costigan-Eaves, P. A history of graphs. Unpublished paper. New Brunswick, New Jersey: Rutgers University Reading Center, 1983.

Craver, J. S. Graph paper from your copier. Tuscon, AZ: H.P. Books, 1980.

Doblin, J. A structure for nontextual communications. In Kolers, Worlstad, & Bouma (Eds.), Processing visible language, 2. New York & London: Plenum Press, 1980.

Funkhouser, H. G. Historical development of the graphical representation of statistical data. Osiris, 1937, 3, 269-404.

Levin, J. R. On functions of pictures in prose. In W. C. Wittrock and F. V. Pirozzolot (Eds.), Neuropsychological and cognitive processes in reading. New York: Academic Press, 1981.

McDonald-Ross, M. Graphics in text. Review of Research in Education, 1977, Vol. 5.

Meyers, C. H. Handbook of basic graphs: A modern approach. Belmont, CA: Dickinson Publishing Co., 1970.

Paller, A., Szoka, K., & Nelson, N. Choosing the right chart, a comprehensive guide for computer graphics users. San Diego, CA: Integrated Software Systems, Inc. (ISSCO), 1981.

Samuels, S. J., LaBerge, D. Toward a theory of automatic information processing in reading. Cognitive Psychology, 1974, 6, 293-323.

Schmid, C. F. The role of standards in graphic presentation. Graphic presentation of statistical information. U.S. Census Bureau Technical Paper No. 43. Washington, DC: U. S. Dept. of Commerce, 1978.

Schmid, C. F., & Schmid, S. E. Handbook of graphic presentation (2nd ed). New York: Ronald Press, J. Wiley & Sons, 1979.

Standards of statistical presentation. Department of the Army Pamphlet 325-10. Washington, DC: April, 1966.

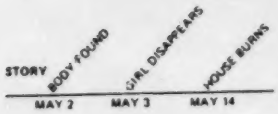
Tufte, E. A. The visual display of quantitative information. Cheshire, Conn.: Graphics Press (Box 430), 1983.

Twyman, M. A schema for the study of graphic language (Tutorial Paper). In Kolers, Worlstad & Bouma (Eds.), Processing of visible language, 1. New York and London: Plenum Press, 1966.

Wainer, H., & Thissen, D. Graphical data analysis. Annual Review of Psychology, 1981, 32, 191-241.

APPENDIX

A TAXONOMY OF GRAPHS



I. LINEAL GRAPHS

Lineal graphs tend to be one dimensional; that is, they are basically a line or a series of lines and it is their characteristic of linearity which transmits the information.

A. SIMPLE LINEAL no branching

1. One Scale (ex. Thermometer)

2. Two Scale (ex. Traditional time line)

Note: Multiple Time Lines with three or more scales are classified with Status Charts in Section IID.

B. COMPLEX LINEAL These have branching, decision points, sub-routines, and/or feedback loops. The important parts of most Complex Lineal Graphs are the connecting lines and junctures (nodes). Line length and position is not important. Lines show connections and sequence, usually not location on coordinates. However some flow charts do have levels and nodes have implied scale position. See section C which follows.

1. Flow

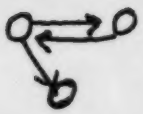
A free form branching pattern information flows in one direction, can have subsystems.

(ex. computer program, wiring diagram, following directions)
(o.t. circut diagram)
(r.t. network, paradigm, algorithm)

Footnote: The term graph and chart are used interchangeably by some authors and not others. This taxonomy favors the term graph but uses the term chart. ex.=example; o.t.=other term, synonym; r.t.=related term, not quite synonym. If this taxonomy is used for classifying graphs it is important to realize that the elements are not mutually exclusive. Any graph may use multiple classification factors.

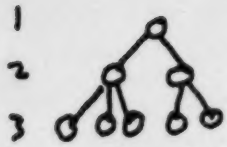
2. Interaction

Information flows in more than one direction.



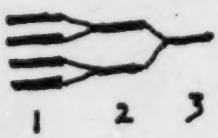
(ex. sociogram, manova)
(r.t. network, model, path, flow, bonding, communications network, simplex)

C. Stratified Lineal Nodes in one plane form a class, level, rank, or generation.



1. Vertical

(ex. organization chart)
(o.t. family tree, organizational structure)
(ex. geneology chart)



2. Horizontal

(ex. sports elimination)

II. RECTANGULAR GRAPHS

These are the typical graphs and they are most often used to show numerical relationships though increasingly nominal data is used. These are the most common graphs and are all based on Cartesian Coordinates. A convention is that time increases along the X axes and these are called "time series" graphs.

(r.t. Statistical Graphs or Charts)

A. CURVE GRAPHS

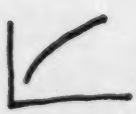
(o.t. line graph or chart)
(r.t. frequency polygon)

1. Line Type

a. Slope

Straight line connects data points

b. Smooth Curve



This may mean smoothed or averaged data. (Nomographs are curves which represent equations or values to be read)

c. Step

May mean averaged or grouped data

d. Spline

Slop line with angles rounded.

e. Ogive

Cumulative frequency distribution

2. Multiple Lines

Good for comparisons, but can be confusing

3. Surface

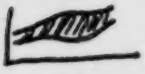
Like a line graph except area below line is shaded.

(o.t. stratum chart)

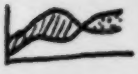
a. Single Surface

b. Multiple Surface

c. Band



Only area between lines is shaded to contrast difference

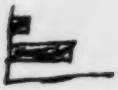


d. Deviation Band

Shows when band difference changes from plus to minus

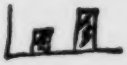
B. BAR GRAPHS

1. Direction of Bar



a. Horizontal Bar

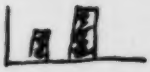
b. Vertical Bar



(o.t. Column Chart or Graph)

The same information can be displayed with Horizontal or Vertical Bar, it is a matter of convention or author's choice

2. Grouping of Bars



a. Simple Bar

Single bar or separated bars

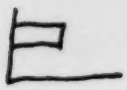
b. Connected Bar

(o.t. Histogram or attached)
In a histogram, bar area is X times Y.

c. Grouped Bars

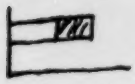
(o.t. Cluster bar)

3. Subdivided Bar



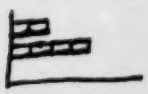
a. Solid Bar

The most common



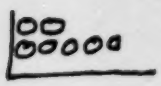
b. Subdivided Bar

(o.t. Segmented Bar)



c. Unit Bar

Shows regular interval units

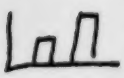


d. Pictograph Bar

Units shown as small pictures or symbols

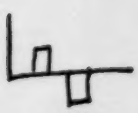
4. Quadrants Used

Positive and/or negative scaling used



a. Positive quadrant used

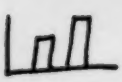
The most common kind (+X and +Y)



b. Negative quadrant(s) used

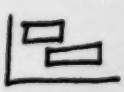
(r.t. deviation bar, sliding bar)

5. Anchoring of Bars



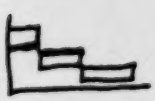
a. Anchored Bar(s)

The most common kind, bar touches base line



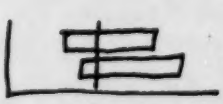
b. Floating Bar

Often used to show range



c. Progressive Bar

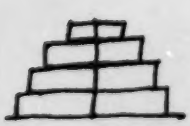
A special type of floating bar in which segments are successive. (o.t. Step by Step Bar)



d. Deviation Bar

Bar crosses dividing line

(o.t. sliding bar)

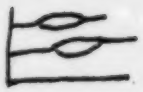


e. Paired Bar

Compares two sets of data on the same horizontal data
ex. Population pyramid, age pyramid
(r.t. bilateral histogram)

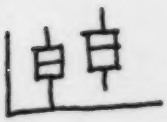
6. Bar Variations

a. Bulge Bar



Used to indicate less exact data (ex. Influence of a king during prime of reign)

b. Box and Whisker Plot



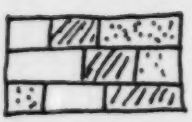
A way of reporting frequency distribution Box shows inter quartile range, tick mark shows mean, whiskers show range.

C. SCATTERGRAMS

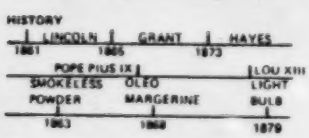
Show many discrete scores or data points more accurate than smoothed curve but harder to generalize or see trends

(o.t. Scatterplots)

D. STATUS GRAPH

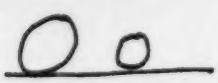


Similar to a Unit Bar Graph but scaling and data are often nominal. It is similar to a common scattergram in that data are given meaning by location in XY coordinates.



(o.t. Status Chart)
(r.t. Multiple Time Line)

E. AREA GRAPHS



These graphs tend to show amount by area (or worse volume, 3-D) They are not recommended because they are difficult to read accurately. A volume scale unit could be calculated, but it is so confusing it defeats the purpose of a graph-visual understanding.

III. CIRCULAR BASE GRAPHS

These graphs are based on polar coordinates

A. PIE GRAPH

1. Simple Pie Graph

Usually used to show percent

(o.t. circle graph)
(r.t. exploded pie)

2. Segmented Pie Graph

Two or more levels (circular base lines)

B. CIRCULAR BAR

1. Star Display

Shows multivariate data
Axes like a circular base bar graph
Useful in showing cyclical data like average rainfall. Hard to read.
Useful for some comparisons

(o.t. Glyph)

2. Wind Rose

Arrows show wind direction, length indicates frequency, feathers indicate force.

C. DIALS

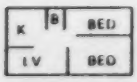
Quicker to read approximate data than digital (symbolic) displays.
Tend to give analogue (continuous) information by location of indicator. May have double scale (see Linear Graphs) for example a spedometer may have kilometer and miles per hour

(ex. Clock face, speedometer, compass rose)

IV. MAPS

All maps show area

A. SINGLE PLANE

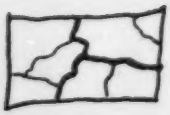


1. Simple area

Uses lines only as boundaries.

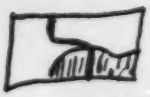
(ex. floor plan, map of states, Venn Diagram)

2. Subdivisions



Areas within areas

(ex. counties within state)
(r.t. maps with shaded or colored areas are "choropleth's" see color note at end)
(r.t. data maps)

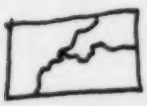


3. Overlapping areas

(ex. water district in part of two counties)

4. Points

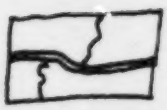
(ex. Location of city)



5. Lines

Lines usually indicate a fixed series of points

(ex. roads)



6. Bands

Lines of two dimensions or thickness. These do not designate area but may indicate intensity or amount of flow.

(ex. busy highways have thicker lines, large army movement)

B. MULTI PLANE

Data from two or more distinct planes shown.

(ex. weather map shows area of states and air condition such as temperature, snow, wind, etc.)

These maps may contain isolines such as isotherms (connecting equal temperature points)

C. TOPOGRAPHIC

Shows height and depth

(ex. Topographic Maps of Mountains, Marine Navigation Chart)
3-D Maps with perspective are classed as pictures

(r.t. contour lines)

D. SPHERICAL

There are some problems of projecting (mapping) a round surface on a flat surface. Cartographers have special coordinates that attempt to solve the problems like Meriator, cylindrical, cone, and polar projections. A thorough classification of these is beyond the present taxonomy.

Note: Maps frequently have much other data superimposed on them such as notes or other graphs. ex. region maps with a pie chart in each region, population dots, pictures. Give such maps multiple classifications.

V. PICTORIAL GRAPHS

These graphs vary from visual reality to abstract Most pictures have actual or implied 3-D prospective scaling.

A. REALISTIC

Maximum data and no apparent distortion

1. Photographs

   Made by a camera or mechanical means

2. Visually Realistic Drawings

   Made by an artist

B. SEMIPICTORIAL

Modest data missing and/or modest distortion of shape (irregular scale)

1. Pictures

   (ex. cartoons)

   a. Generic

      (ex. a man meant to be any man)

   b. Specific

      (ex. Popeye)

2. Schematic Drawings

   Tend to fundamental or internal elements

   a. Exploded

      (ex. parts of carburetor laid out in order of assembly)

   b. Cut-away

      (ex. view of engine cut in half)

   c. Simplified elements

      (ex. chemical plant units simplified; framework of house)

      (r.t. Isometric drawings)

3. Silhouette

C. ABSTRACT PICTORIAL

1. Data amount

a. Data Missing

(ex. fuzzy photograph)

b. Data Added

(unrealistically)

2. Distorted shape

May have irregular scale and/or data misplaced

D. DATA DISPLAYS IN 3 DIMENSIONS

(ex. Bar chart using X, Y or Z axis)

non pictorial, no visual reality

VI. COLOR AND SCALING, AND OTHER FACTORS

This section of the taxonomy is outside the theory and basic framework of graphs however it brings in important factors used in reading and writing graphs.

A. COLOR

1. One color (in addition to background)

2. Two colors

3. Three colors

4. More than 3 colors

(ex. full color photograph)

B. SHADING

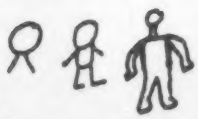
1. One shading or basic ink color used

2. Two shadings used

3. Three or more shadings used

C. SERIES

Series refers here to a sequence of single graphs that show changes over time or some other variable.



(ex. series of cartoon pictures, sequence of pictures showing tadpole development)

C. SERIES (continued)

Note: The term "time series" is used by some authors in a different sense - "time series" can mean a typical rectangular graph such as a curve in which the variable "time" increases along the x axis.

D. FALSE 3D

False 3D is supposed to make the graph more interesting by adding a third dimension to bars, pie charts or maps. The third dimension adds no new information, Tufte calls it "chart junk".

(o.t. Decorative 3D, 2 1/2D)

E. SYMBOLS

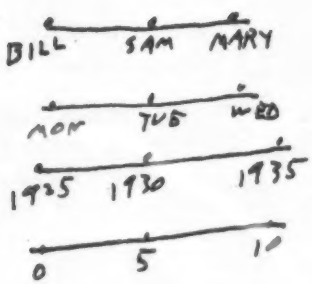
This is an expandable catagory which varies widely with the field or subject matter of the graph.

1. Arrows

   Widely used to show motion or directionality

2. Other

F. SCALING



1. Nominal

2. Ordinal

3. Interval

4. Ratio

G. SCALING UNITS

1. Arithmetic (Decimal)

   a. Raw Data

      (ex. Raw Scores)

   b. Percentage

   c. Derived Scores or data

      (ex. IQ, Standard Scores, Grade Equivalent, Percentile)

   d. Averaged

      (ex. moving average)

2. Time Units

(ex. hours, minutes, seconds)

3. Degree Units

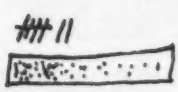
(ex. degrees, minutes, seconds)

4. Logarithmic Units

5. Perspective Units

6. Other Units

(ex. square roots, radii)



D. TALLY SYSTEMS

Show amount in units, pen strokes, dots, shaded squares, etc.

(r.t. Graphical Rational Patterns, Chernoff's Faces)

Fig. 1. This figure illustrates that the commonly used Cartesian or rectangular coordinates can show negative amounts with a bar chart. Dashed line and heavy dot on top of bar show that the same information can be used to form a curve graph or points on a scattergram (dots).

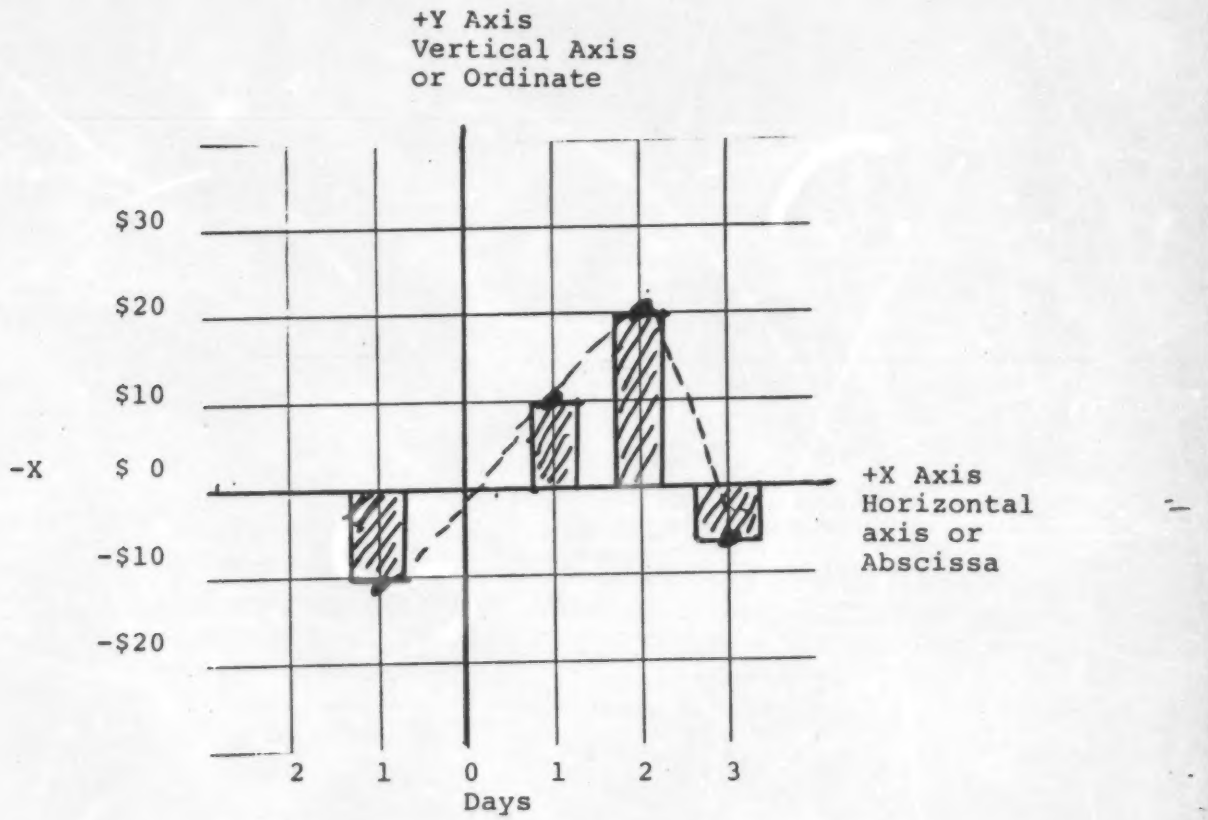


Fig. 2. A polar coordinate system with base ring scaled in percent. This is the coordinate system underlying all pie charts. Rings and axes can have any scaling.

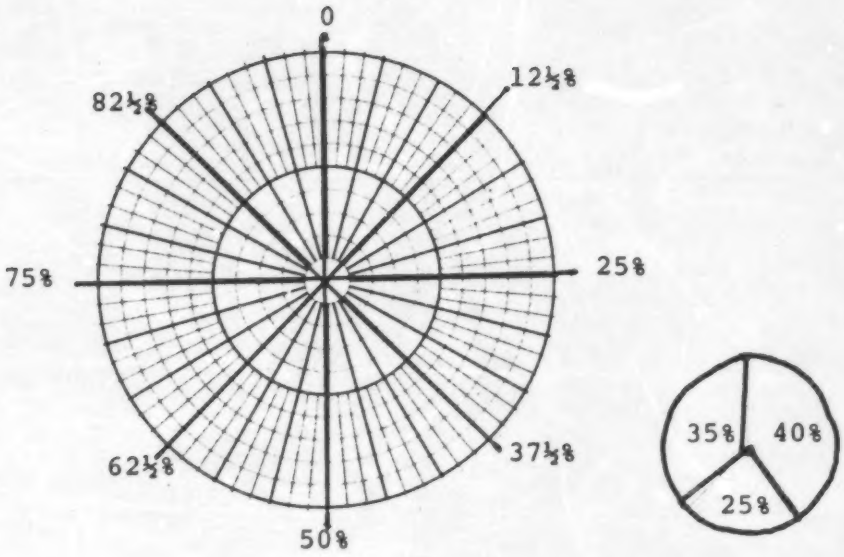
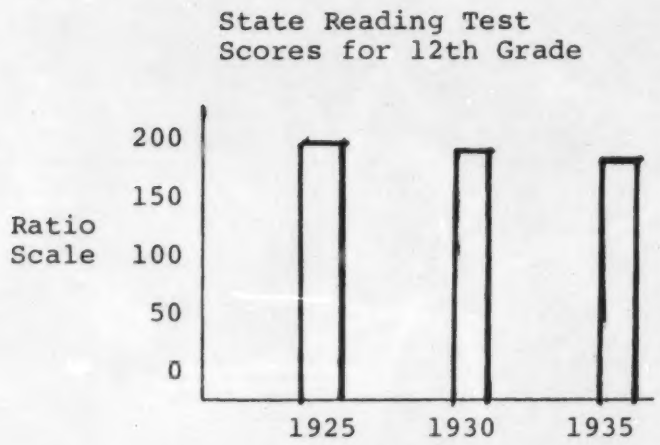


Fig. 3. The same data plotted on an interval scale and a ratio scale (Y axis) of a bar graph. The ratio scale shows that compared to total score there is very little change while the interval scale starting at 150 magnifies the change and it looks like reading scores are descending at an alarming rate.



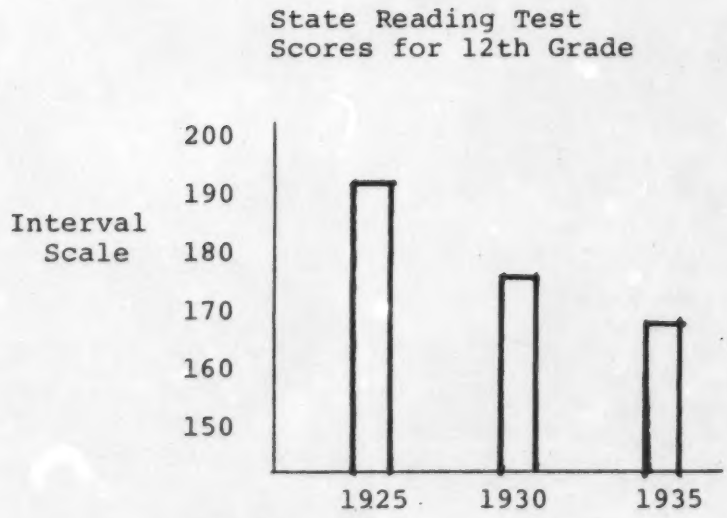
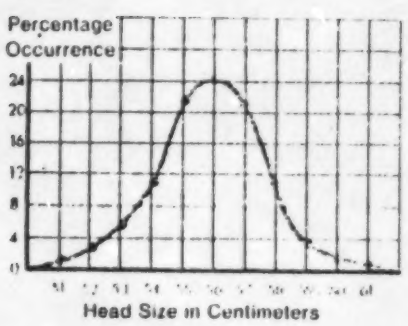
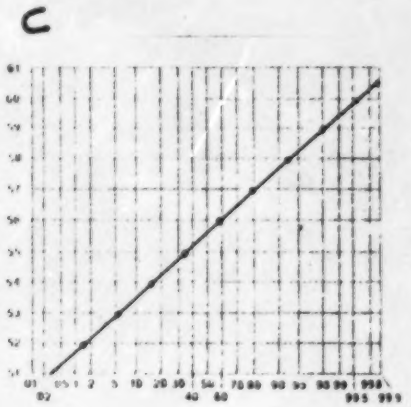


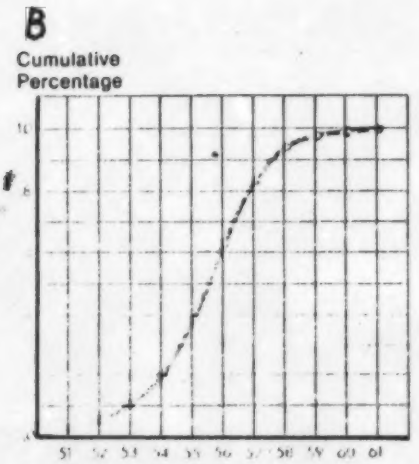
Fig. 4. Different ways of showing normally distributed data.

Illustration A shows data normally distributed. Illustration B shows the same data with a cumulative percentage which in this instance is like a percentile. Illustration C shows the same data but the bottom scale is a probability scale. Illustration D shows how an extended line can estimate points on a normal curve or how you can estimate probability or percentile given the assumption of normally distributed data. Graph from Carver (1980).

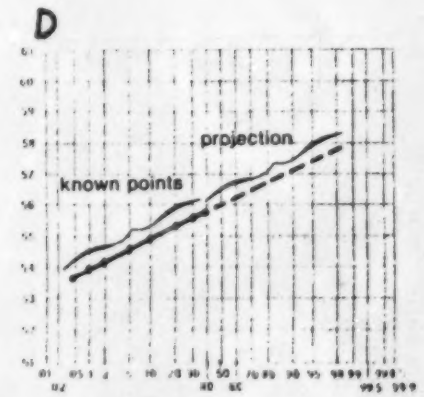




Probability grid with percentage of head size plotted



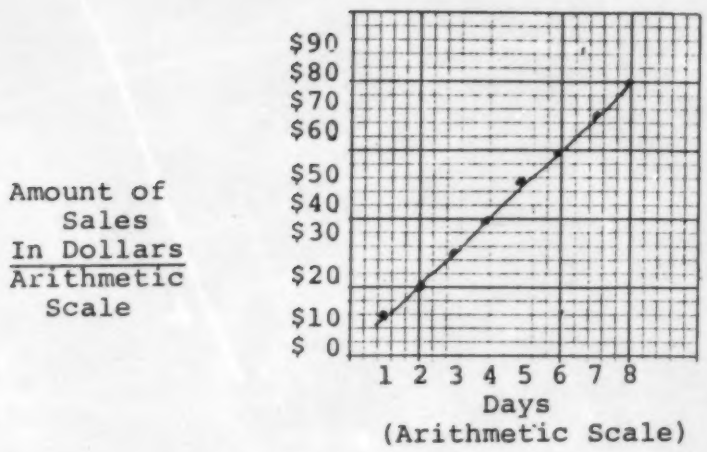
Ogive Curve

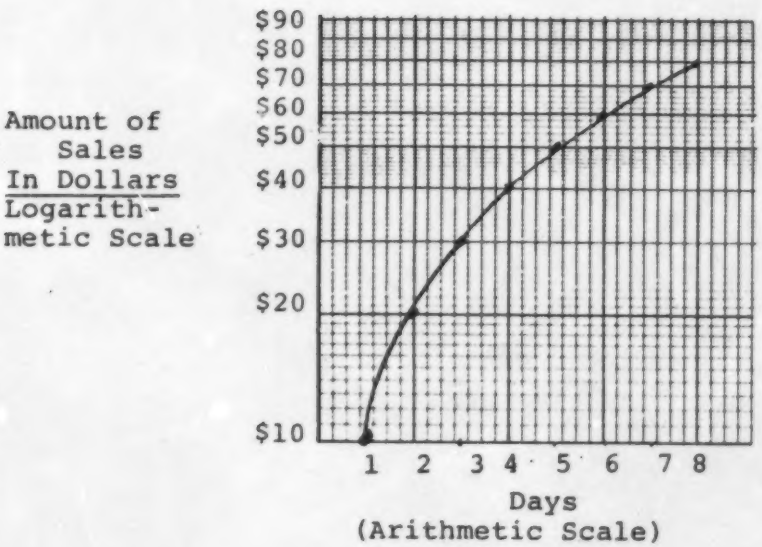


Projection of known data on probability grid to predict occurrences

Fig. 5. Arithmetic and Semi-log graphs showing the same data.

Both graphs show an increase of sales at $10 per day. If we look at the curve on the top scale it looks like there is no end of growth in sight for Maria's Lemonade Stand, but seen as a percentage of sales in the logarithmic scale, the curve, while still showing growth, indicates that Maria might not need to expand quite as soon, or an investor who expected to buy in on day 7, might not double his money too quickly.





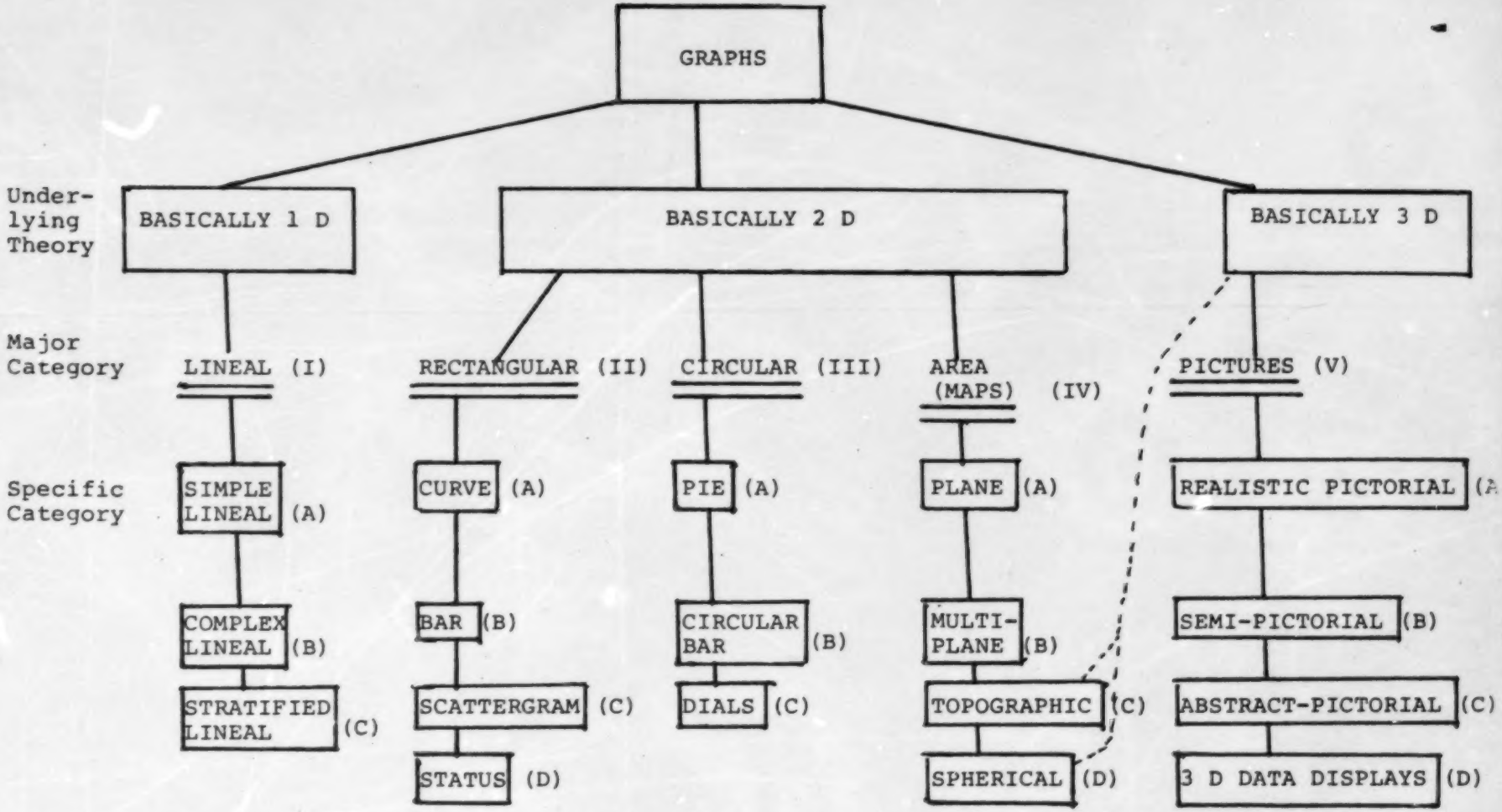


Fig. 6. Outline of a taxonomy of graphs. Each of these categories is expanded and illustrated in the appendix.